

10/528675

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 10/528,675

Source: P4/10

Date Processed by STIC: 4/4/05

ENTERED



PCT

RAW SEQUENCE LISTING

DATE: 04/04/2005

PATENT APPLICATION: US/10/528,675

TIME: 10:59:39

Input Set : A:\0185653.ST25.txt

Output Set: N:\CRF4\04042005\J528675.raw

3 <110> APPLICANT: Hoshino , Tatsuo
 4 Miyazaki, Taro
 5 Sugisawa, Teruhide
 7 <120> TITLE OF INVENTION: Aldehyde Dehydrogenase Gene
 9 <130> FILE REFERENCE: C038435/0185653
 C--> 11 <140> CURRENT APPLICATION NUMBER: US/10/528,675
 C--> 11 <141> CURRENT FILING DATE: 2005-03-23
 11 <160> NUMBER OF SEQ ID NOS: 9
 13 <170> SOFTWARE: PatentIn version 3.3
 15 <210> SEQ ID NO: 1
 16 <211> LENGTH: 3408
 17 <212> TYPE: DNA
 18 <213> ORGANISM: Gluconobacter oxydans
 20 <400> SEQUENCE: 1
 21 gcgactggca gcagcgcaac tatgaccact atggcctgcc gccctattgg atctaactga 60
 23 tccagtaagc caccatcagc cggcccctgc gggggccggc tttttgcgct agaccccgcc 120
 25 gaggtgctgt cgtaacctaa ggtcacatct ttacttccac atccgccctt gtcagttctg 180
 27 acgtgacaaa ttgtcgcggt catgctgctg aatgcggatg ccagtcccag atccaagccc 240
 29 gacgcaagga gacgtagatg ttacccaaat cattgaaaca taagaatggc gccatgcgcc 300
 31 ttgtcgcagc ctgcaccctt gcgctgatga tggcgcgggg tgcccattgc caggtaaacc 360
 33 cggtcgaagt gccgggtggc gcgaacgaga cctttacctc gcgcgtgctg accaccggcc 420
 35 tgtcgaaccc ttgggaaatc acctggggcc cgcacaatat gctgtgggtg accgagcgat 480
 37 cttccggcga agtgacgcgc gtgcacccca ataccggcg gcagcaggtc ctgctgaccc 540
 39 tgaccgatatt cagcgctgat gtgcaacacc agggcctact tggcctcgcg ctgcatcctg 600
 41 agtttatgca agagagcggc aacgactacg tctatatcgt ctacacttat aacaccggca 660
 43 ccgaagaagc gcccgatccg catcaaaagc tgggtgcgta tgccatgac gctgccgcgc 720
 45 agcagctggt cgatccggtt gatctggtcg caggcattcc cgcaggcaac gaccacaatg 780
 47 gcggtcgcat caaatcgcgc cccgatggcc aacacatctt ttacacgctg ggcgagcaag 840
 49 gcgcgaactt tggcggtaac ttccgcgcgc cgaaccacgc gcaactgctg ccgacgcaag 900
 51 agcaggtcga cgcgggcgat tgggtgcctt attcgggcaa gatcctgcgc gtgaaccttg 960
 53 acggcacgat ccccgaaagc aaccccgaga tcgagggcgt gcgtagccat atctttacct 1020
 55 atggccaccg taaccgcgag ggcatacact ttggccccga cggcaccatt tatgccaccg 1080
 57 aacacggccc cgatacggat gacgagctga acatcatcgc cggcggtggc aactatgggt 1140
 59 ggccgaatgt ggccggctat cgcgatggca aatcctatgt ctacgctgat tggagccaag 1200
 61 cgcccgtgta ccagcgttac accggtcgcg ccggtatccc cgacaccgtg ccgcaattcc 1260
 63 ccgagctgga attcgcgcgc gagatggtcg atccgctgac aacctattgg acggtggata 1320
 65 atgattacga tttcaccgcc aattgcggct ggatctgtaa tccgacgac gcgccttcgt 1380
 67 ctgcctatta ctatgcggcg ggcgagagcg gtatcgcggc ttgggataat tcgatcctga 1440
 69 tcccgacgct gaaacatggc ggcatactat tgcagcacct cagcgatgat ggccaatctg 1500
 71 tcgacggcct gcccgagctg tggttcagca cccagaaccg ctatcgcgat atcgagatca 1560
 73 gccccgataa ccatgttttt gtggcgaccg acaactttgg cacctcggcg cagaaatatg 1620
 75 gcgagaccgg ctttaccaac gtgctgcata accccggcgc gatccttgct tttagctatg 1680
 77 tcggcgagga tgctgcgggt cagaccggaa tgatgaccgc gcccgaccg cagacgcaat 1740

p.6

RAW SEQUENCE LISTING

DATE: 04/04/2005

PATENT APPLICATION: US/10/528,675

TIME: 10:59:39

Input Set : A:\0185653.ST25.txt

Output Set: N:\CRF4\04042005\J528675.raw

```

79 acacgcgaagt gcccgcgcgag ggtgcaggcg cgggcgcgcac tgagggttgcg gatgtcgatt 1800
81 acgacacgct gttcaccgaa ggccagaccc tttatggcag cgcattgtgcc gcgtgccatg 1860
83 gtgccgctgg ccaagggtgcg cagggcccgga cctttgtggg cgtgccggat gtgacgggtg 1920
85 acaaggacta ccttgcccgc accatcatcc acggttttgg ctatatgccg tcgtttgcga 1980
87 ctcggtctgga tgacgaggag gttgccgcca tcgcgacctt tatccgcaac agctggggca 2040
89 atgacgaagg catcctgacc ccggccgagg ccgctgccac ccgctgaatg ctgtaaaaaac 2100
91 caccctcgcc tgcacatcag gcgggggtat ttcattttatt ttcacatctg cctttgacat 2160
93 gtgccgctat cacggttaat gcggcccttc ggctgtttctg ggtctaagcg ggtgtgttgc 2220
95 ccgataagag agacggttca gtccctcccg ccctatttag ggcccattta ggcagaatag 2280
97 ttttgactca tcaaaatatc gccgcgcctc tggccgcggc cctttcgcaa cgtggatatg 2340
99 aaacgctgac cgcctgtcag caagctgtgc ttgcgcccga ggctgatggc cgcgacctgc 2400
101 tgggtgtcggc acagaccggt tcgggtaaga cgggtggcctt tggatatgca gtcgcgcccg 2460
103 accttttggg cgacgacaat atcctgccgc tgaacacgcc gcctgttgcg ctgttcatcg 2520
105 ccccccacgcg cgagcttgcg ctgcaagttg ctcaggaact gacctggctt tacgccaatg 2580
107 caggtgcccc gatcgcgcacc tgcgtcggcg gtatggatta ccgcaccgag cgcgcgcccc 2640
109 ttgcacgtct gccgcaaate gttgtcggca cgcccggcgg tctgcgcgac catatcgacc 2700
111 gtggcgccct tgacctgtcc gaattgcgcg tgacctgtct ggacgaagcg gatgagatgc 2760
113 tcgacctcgg cttccgcgat gatctgcaat atatcttgca agccgcgccc gaagatcgcc 2820
115 gcacgctgat gttctcggcc accgtgcgcg gcgagattga aaaactggcc cgcgacttcc 2880
117 aaaatgacgc cctgcgtctg gaaacccgtg gcgaggccaa gcagcacaac gacatcagct 2940
119 accaagcttt gtcggtcacc atgcgcgac gcgaaaacgc cattttcaac atgctgcgtt 3000
121 tttatgaate gcgcacggcg atcatcttct gcaagaccgc cgccaatgtg aatgatctgc 3060
123 tgtcgcggat gacgggtcgt ggcttccgcg tgggtggcct gtcggggcgag ctgtcgcaac 3120
125 aggaacgcac caacgcgctg caagcgtgc gtgatggcg cgccaacggt tgtatcgca 3180
127 ccgacgtcgc ggcgcgcgcc attgacttgc cgggcctcga gctggtgatc cactacgatc 3240
129 tgccgaccaa tgccgaaacc ctgctgcacc gctcggggcg taccggccgc cgggtgccaa 3300
131 gggcgctctc gcgctgatcg tcacccccgg cgatttcaaa aaagcgcagc gtttgctgag 3360
133 ctttgccaaa gtgaccgcgg aatggggcaa ggcgccttcg gccgaaga 3408

```

136 <210> SEQ ID NO: 2

137 <211> LENGTH: 609

138 <212> TYPE: PRT

139 <213> ORGANISM: Gluconobacter oxydans

141 <400> SEQUENCE: 2

```

143 Met Leu Pro Lys Ser Leu Lys His Lys Asn Gly Ala Met Arg Leu Val
144 1 5 10 15
147 Ala Ala Ser Thr Leu Ala Leu Met Ile Gly Ala Gly Ala His Ala Gln
148 20 25 30
151 Val Asn Pro Val Glu Val Pro Val Gly Ala Asn Glu Thr Phe Thr Ser
152 35 40 45
155 Arg Val Leu Thr Thr Gly Leu Ser Asn Pro Trp Glu Ile Thr Trp Gly
156 50 55 60
159 Pro Asp Asn Met Leu Trp Val Thr Glu Arg Ser Ser Gly Glu Val Thr
160 65 70 75 80
163 Arg Val Asp Pro Asn Thr Gly Glu Gln Gln Val Leu Leu Thr Leu Thr
164 85 90 95
167 Asp Phe Ser Val Asp Val Gln His Gln Gly Leu Leu Gly Leu Ala Leu
168 100 105 110
171 His Pro Glu Phe Met Gln Glu Ser Gly Asn Asp Tyr Val Tyr Ile Val
172 115 120 125

```

RAW SEQUENCE LISTING

DATE: 04/04/2005

PATENT APPLICATION: US/10/528,675

TIME: 10:59:39

Input Set : A:\0185653.ST25.txt

Output Set: N:\CRF4\04042005\J528675.raw

```

175 Tyr Thr Tyr Asn Thr Gly Thr Glu Glu Ala Pro Asp Pro His Gln Lys
176      130      135      140
179 Leu Val Arg Tyr Ala Tyr Asp Ala Ala Ala Gln Gln Leu Val Asp Pro
180 145      150      155      160
183 Val Asp Leu Val Ala Gly Ile Pro Ala Gly Asn Asp His Asn Gly Gly
184      165      170      175
187 Arg Ile Lys Phe Ala Pro Asp Gly Gln His Ile Phe Tyr Thr Leu Gly
188      180      185      190
191 Glu Gln Gly Ala Asn Phe Gly Gly Asn Phe Arg Arg Pro Asn His Ala
192      195      200      205
195 Gln Leu Leu Pro Thr Gln Glu Gln Val Asp Ala Gly Asp Trp Val Ala
196      210      215      220
199 Tyr Ser Gly Lys Ile Leu Arg Val Asn Leu Asp Gly Thr Ile Pro Glu
200 225      230      235      240
203 Asp Asn Pro Glu Ile Glu Gly Val Arg Ser His Ile Phe Thr Tyr Gly
204      245      250      255
207 His Arg Asn Pro Gln Gly Ile Thr Phe Gly Pro Asp Gly Thr Ile Tyr
208      260      265      270
211 Ala Thr Glu His Gly Pro Asp Thr Asp Asp Glu Leu Asn Ile Ile Ala
212      275      280      285
215 Gly Gly Gly Asn Tyr Gly Trp Pro Asn Val Ala Gly Tyr Arg Asp Gly
216      290      295      300
219 Lys Ser Tyr Val Tyr Ala Asp Trp Ser Gln Ala Pro Ala Asp Gln Arg
220 305      310      315      320
223 Tyr Thr Gly Arg Ala Gly Ile Pro Asp Thr Val Pro Gln Phe Pro Glu
224      325      330      335
227 Leu Glu Phe Ala Pro Glu Met Val Asp Pro Leu Thr Thr Tyr Trp Thr
228      340      345      350
231 Val Asp Asn Asp Tyr Asp Phe Thr Ala Asn Cys Gly Trp Ile Cys Asn
232      355      360      365
235 Pro Thr Ile Ala Pro Ser Ser Ala Tyr Tyr Tyr Ala Ala Gly Glu Ser
236      370      375      380
239 Gly Ile Ala Ala Trp Asp Asn Ser Ile Leu Ile Pro Thr Leu Lys His
240 385      390      395      400
243 Gly Gly Ile Tyr Val Gln His Leu Ser Asp Asp Gly Gln Ser Val Asp
244      405      410      415
247 Gly Leu Pro Glu Leu Trp Phe Ser Thr Gln Asn Arg Tyr Arg Asp Ile
248      420      425      430
251 Glu Ile Ser Pro Asp Asn His Val Phe Val Ala Thr Asp Asn Phe Gly
252      435      440      445
255 Thr Ser Ala Gln Lys Tyr Gly Glu Thr Gly Phe Thr Asn Val Leu His
256      450      455      460
259 Asn Pro Gly Ala Ile Leu Val Phe Ser Tyr Val Gly Glu Asp Ala Ala
260 465      470      475      480
263 Gly Gln Thr Gly Met Met Thr Ala Pro Ala Pro Gln Thr Gln Tyr Thr
264      485      490      495
267 Gln Val Pro Ala Glu Gly Ala Gly Ala Gly Ala Thr Glu Val Ala Asp
268      500      505      510
271 Val Asp Tyr Asp Thr Leu Phe Thr Glu Gly Gln Thr Leu Tyr Gly Ser

```

RAW SEQUENCE LISTING

DATE: 04/04/2005

PATENT APPLICATION: US/10/528,675

TIME: 10:59:39

Input Set : A:\0185653.ST25.txt

Output Set: N:\CRF4\04042005\J528675.raw

```

272          515          520          525
275 Ala Cys Ala Ala Cys His Gly Ala Ala Gly Gln Gly Ala Gln Gly Pro
276          530          535          540
279 Thr Phe Val Gly Val Pro Asp Val Thr Gly Asp Lys Asp Tyr Leu Ala
280 545          550          555          560
283 Arg Thr Ile Ile His Gly Phe Gly Tyr Met Pro Ser Phe Ala Thr Arg
284          565          570          575
287 Leu Asp Asp Glu Glu Val Ala Ala Ile Ala Thr Phe Ile Arg Asn Ser
288          580          585          590
291 Trp Gly Asn Asp Glu Gly Ile Leu Thr Pro Ala Glu Ala Ala Ala Thr
292          595          600          605
295 Arg
299 <210> SEQ ID NO: 3
300 <211> LENGTH: 14
301 <212> TYPE: PRT
302 <213> ORGANISM: Gluconobacter oxydans
305 <220> FEATURE:
W--> 306 <221> NAME/KEY: X
307 <222> LOCATION: (2)..(2)
308 <223> OTHER INFORMATION: X can be any naturally occurring amino acid.
310 <220> FEATURE:
W--> 311 <221> NAME/KEY: X
312 <222> LOCATION: (4)..(4)
313 <223> OTHER INFORMATION: X is P or K.
315 <400> SEQUENCE: 3
W--> 317 Gln Xaa Asn Xaa Val Glu Val Pro Val Gly Ala Asn Glu Thr
318 1          5          10
321 <210> SEQ ID NO: 4
322 <211> LENGTH: 31
323 <212> TYPE: PRT
324 <213> ORGANISM: Gluconobacter oxydans
326 <400> SEQUENCE: 4
328 Met Leu Pro Lys Ser Leu Lys His Lys Asn Gly Ala Met Arg Leu Val
329 1          5          10          15
332 Ala Ala Ser Thr Leu Ala Leu Met Ile Gly Ala Gly Ala His Ala
333          20          25          30
336 <210> SEQ ID NO: 5
337 <211> LENGTH: 578
338 <212> TYPE: PRT
339 <213> ORGANISM: Gluconobacter oxydans
341 <400> SEQUENCE: 5
343 Gln Val Asn Pro Val Glu Val Pro Val Gly Ala Asn Glu Thr Phe Thr
344 1          5          10          15
347 Ser Arg Val Leu Thr Thr Gly Leu Ser Asn Pro Trp Glu Ile Thr Trp
348          20          25          30
351 Gly Pro Asp Asn Met Leu Trp Val Thr Glu Arg Ser Ser Gly Glu Val
352          35          40          45
355 Thr Arg Val Asp Pro Asn Thr Gly Glu Gln Gln Val Leu Leu Thr Leu
356          50          55          60

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/528,675

DATE: 04/04/2005

TIME: 10:59:39

Input Set : A:\0185653.ST25.txt

Output Set: N:\CRF4\04042005\J528675.raw

```

359 Thr Asp Phe Ser Val Asp Val Gln His Gln Gly Leu Leu Gly Leu Ala
360 65              70              75              80
363 Leu His Pro Glu Phe Met Gln Glu Ser Gly Asn Asp Tyr Val Tyr Ile
364              85              90              95
367 Val Tyr Thr Tyr Asn Thr Gly Thr Glu Glu Ala Pro Asp Pro His Gln
368              100             105             110
371 Lys Leu Val Arg Tyr Ala Tyr Asp Ala Ala Ala Gln Gln Leu Val Asp
372              115             120             125
375 Pro Val Asp Leu Val Ala Gly Ile Pro Ala Gly Asn Asp His Asn Gly
376              130             135             140
379 Gly Arg Ile Lys Phe Ala Pro Asp Gly Gln His Ile Phe Tyr Thr Leu
380 145              150             155             160
383 Gly Glu Gln Gly Ala Asn Phe Gly Gly Asn Phe Arg Arg Pro Asn His
384              165             170             175
387 Ala Gln Leu Leu Pro Thr Gln Glu Gln Val Asp Ala Gly Asp Trp Val
388              180             185             190
391 Ala Tyr Ser Gly Lys Ile Leu Arg Val Asn Leu Asp Gly Thr Ile Pro
392              195             200             205
395 Glu Asp Asn Pro Glu Ile Glu Gly Val Arg Ser His Ile Phe Thr Tyr
396              210             215             220
399 Gly His Arg Asn Pro Gln Gly Ile Thr Phe Gly Pro Asp Gly Thr Ile
400 225              230             235             240
403 Tyr Ala Thr Glu His Gly Pro Asp Thr Asp Asp Glu Leu Asn Ile Ile
404              245             250             255
407 Ala Gly Gly Gly Asn Tyr Gly Trp Pro Asn Val Ala Gly Tyr Arg Asp
408              260             265             270
411 Gly Lys Ser Tyr Val Tyr Ala Asp Trp Ser Gln Ala Pro Ala Asp Gln
412              275             280             285
415 Arg Tyr Thr Gly Arg Ala Gly Ile Pro Asp Thr Val Pro Gln Phe Pro
416              290             295             300
419 Glu Leu Glu Phe Ala Pro Glu Met Val Asp Pro Leu Thr Thr Tyr Trp
420 305              310             315             320
423 Thr Val Asp Asn Asp Tyr Asp Phe Thr Ala Asn Cys Gly Trp Ile Cys
424              325             330             335
427 Asn Pro Thr Ile Ala Pro Ser Ser Ala Tyr Tyr Tyr Ala Ala Gly Glu
428              340             345             350
431 Ser Gly Ile Ala Ala Trp Asp Asn Ser Ile Leu Ile Pro Thr Leu Lys
432              355             360             365
435 His Gly Gly Ile Tyr Val Gln His Leu Ser Asp Asp Gly Gln Ser Val
436              370             375             380
439 Asp Gly Leu Pro Glu Leu Trp Phe Ser Thr Gln Asn Arg Tyr Arg Asp
440 385              390             395             400
443 Ile Glu Ile Ser Pro Asp Asn His Val Phe Val Ala Thr Asp Asn Phe
444              405             410             415
447 Gly Thr Ser Ala Gln Lys Tyr Gly Glu Thr Gly Phe Thr Asn Val Leu
448              420             425             430
451 His Asn Pro Gly Ala Ile Leu Val Phe Ser Tyr Val Gly Glu Asp Ala
452              435             440             445
455 Ala Gly Gln Thr Gly Met Met Thr Ala Pro Ala Pro Gln Thr Gln Tyr

```

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/10/528,675

DATE: 04/04/2005
TIME: 10:59:40

Input Set : A:\0185653.ST25.txt
Output Set: N:\CRF4\04042005\J528675.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:3; Xaa Pos. 2/4
Seq#:7; N Pos. 9

Invalid <213> Response:

Use of "Artificial" only as "<213> Organism" response is incomplete, per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:6,7,8,9

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/528,675

DATE: 04/04/2005

TIME: 10:59:40

Input Set : A:\0185653.ST25.txt

Output Set: N:\CRF4\04042005\J528675.raw

L:11 M:270 C: Current Application Number differs, Replaced Current Application No
L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:306 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:3
L:311 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:3
L:317 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3 after pos.:0
L:501 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:6
L:506 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:6
L:511 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:6
L:516 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:6
L:534 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:7
L:539 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:7
L:549 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:7
L:554 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:7
L:559 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:7 after pos.:0